Advanced Materials Research *Online: 2014-01-16*
ISSN: 1662-8985, Vol. 878, pp 797-805
doi:10.4028/www.scientific.net/AMR.878.797

Spatial Distribution of As-concentrations in the Contaminated Site by the Highly Toxic As-Chemicals

Fengwu PENG [1, a], Ting ZHOU[1, b], Shuanjin WEI [1,c], Xuefeng WANG [1,d], Lijuan YOU [1,e] and Ning WANG [1,f]

[1] Institute of Chemical Defense, Beijing, 102205, China

[a] 13910060501@139.com, [b]zhting_12@sina.com(corresponding author), [c]13681472182@139.com
[d]youlijuan8743@163.com, [e]13910718536@139.com, [f]13601357303@139.com

Keywords: Highly toxic As-chemicals; As-contaminated Soil; Geostatistics; Semivariogram; Spatial variability

Abstract. A total of 2668 surface soil samples (-0.5 m ~ 0 m) and 637 sub-layer soil samples (-1.0 m ~ -0.5 m) were collected from a contaminated site with an area of 26.19 hectares and analyzed the concentration of arsenic, so as to get the spatial distribution of arsenic concentrations in the soil contaminated by highly toxic As-chemicals. The results indicated that spatial correlation of arsenic content in surface soil was moderate. The optimal semivariogram model was the exponential model, and the best interpolation model was simple Kriging interpolation. The arsenic content of sub-layer was similar to that of surface soil in space structure, but weakly correlated. Analysis on the arsenic content in vertical sections showed that contamination was mainly concentrated in surface soil. The arsenic content in most of the collected samples reduced as the soil layer depth increased. Spatial variability of arsenic could be significantly affected by anthropogenic factors such as excavation & destruction of highly toxic As-chemicals and farming.

Introduction

Half a century ago, a large lot of discarded shells filled with highly toxic As-chemicals were found and simply destroyed by means of detonation, resulting in arsenic contamination of an area of 26.19 ha. To remediation the area, it is now required to investigate and understand the trend and characteristic of arsenic distribution in the area. In this paper, geostatistics is used in combination with the geographic information system (GIS) technology to analyse and study the spatial structure and distribution of arsenic contamination at the site.

Geostatistical analysis has proved to be one of the most effective methods for analysing spatial distribution characteristics and variation patterns of heavy metals in soil, and it is currently the most widely used methods of spatial analysis. Domestic and overseas literature in this regard is abundant; however, the objects of study are mostly areas with relatively low contamination levels, where the parent materials contribute more or a similar amount of heavy metals in the soil compared to the external contamination sources [1]. A site contaminated by destruction of abandoned shells filled with arsenic-containing high-hazard chemicals is special in terms of contamination degree, contamination distribution and the like and has many unique characteristics different from the objects of previous studies. Meanwhile, most previous studies adopted simple random sampling methods, which are of insufficient representativeness and sampling density for contamination areas varying greatly in stochastic components of variogram values among sampling points and masking of small-scale structures [1]. In this study, the survey area mainly consists of hillsides and valleys, and the topography is relatively complex. Previous studies showed that, in regions with complex terrain, increasing sampling density will relatively sharply improve the interpolation accuracy, and increasing the number of samples can greatly improve the interpolation accuracy [2]. Therefore, in combination with the contamination accident monitoring soil sampling provisions of HJ/T166-2004 <Technical specification for soil environmental monitoring>, this study uses the 10m × 10m distribution grid sampling method to study the distribution of arsenic contamination in the survey area.

Materials and method

Samples collection. The study area is located in mountains on the sunny slope of a range of southeast - northwest mountains and the valley at its foot at latitude 43.59° ~ 43.62° and longitude 128.60° ~ 128.62°, as shown in Figure 1 [3].

Fig.1 Sketch map of the study area and the distribution of soil samples

According to the prior findings of the contamination cause, soil thickness & groundwater level, local farming and crop root depth of the survey area, sampling points were arranged in 10m × 10m distribution grids, the grid intersection points were taken as soil survey sampling points, and the sampling depth was 0.5 m [4]. Sampling points were arranged in 10 m × 10 m distribution grids. Collected soil samples were cut at surface (-0.50 m ~ 0) and sub-layer (-1.00 m~-0.50 m) depths, 2,668 surface samples and 637 sub-layer samples were collected.

Determination of arsenic. The collected soil samples were crumbed, spread on plastic trays and dried naturally. After air-dried, the soil samples were ground with an agate mortar and screened with a 100-mesh sieve. Soil sample test methods: digested with sulfuric acid - nitric acid - perchloric acid, tested by the silver diethyldithiocarbamate method, and then verified by the FI-HG-AAS method. GBW-07403 standard soil samples were used for quality control in the test process.

Data processing. Since the presence of outliers have significant impact on variograms, excluding these outliers before the calculation is very necessary [5]. In the data processing of this study, the domain approach was used to identify outliers, namely, data outside [$\bar{x}$ ±3s] (sample mean $\bar{x}$ plus or minus three times the standard deviation), and then the maximum and minimum normal values were used to replace outliers respectively [1, 6]. After the elimination of outliers, the surface and sub-layer arsenic content values were logarithmically transformed, u tested [7] and found to comply with or approach to the normal distribution. Therefore, in the semi-variance analysis, the arsenic content values of the surface layer and the sub-layer were logarithmically transformed, together with the relative geographic coordinates of the sampling points, input into the statistical software Arcgis 10.0 to fit semivariograms; the optimal fitting model and its parameters were selected, and the arsenic content values were interpolated to obtain the spatial distribution of heavy metal content in the soil of the study area

Results and analysis

Descriptive statistics of As-concentrations in the soil. The statistics results of the original data of measured As-concentrations in the surface layer and the sub-layer and the data after elimination of outliers and logarithmic transformation are as shown in Table 1.

Table 1 Descriptive statistics of soil As-concentrations

Soil layer		Minimum value	Maximum value	Average value	Standard deviation	Average variation coefficient	Kurtosis (g_1)	Kurtosis standard error(σg_1)	Skewness (g_2)	Skewness standard error(σg_2)
Surface layer	Raw data	0.09	758.44	30.28	53.39	1.76	39.10		5.15	
	Logarithmic transformation	-0.11	2.67	1.19	0.51		0.19	0.097	0.046	0.049
Sub-layer	Raw data	2.23	1372.87	52.03	78.85	1.52	136.00		9.40	
	Logarithmic transformation	0.42	2.64	1.52	0.40		-0.045	0.195	-0.081	-0.098

Note: The minimum, maximum and mean values of the original data in the table are all in mg·kg$^{-1.}$

Table 1 showed that, in terms of the skewness coefficient and kurtosis coefficient, the original data was not normally distributed; the variation coefficient of the As-concentration original data of the surface layer and the sub-layer was 1.76 and 1.52 respectively, both being strong variations; with regard to 57.38% of the surface layer samples and 85.56% of the sub-layer samples， the As-concentration exceed the upper limit of the local background value; the mean value of As-concentrations of the surface layer and 85.56% of the sub-layer was 30.28 mg·kg^{-1} and 52.03 mg·kg^{-1} respectively, both higher than the national grade II standard as specified in GB15618-1995 Environmental quality standard for soils. These results indicated that, due to anthropogenic sources of contamination, the heavy metals in certain grid in the study area were significantly increased; what's more, arsenic contamination was hazardous, which might affect agricultural production and human health [8].

As the geostatistical analysis is based on the normal data, the original data must be normalized. After the elimination of outliers, the original data was logarithmically transformed and then analyzed on the Arc Map platform. As shown in Table 1, the formula (1)is used for u test, i.e..

$$u = \frac{|g - 0|}{\sigma_g} < u_{0.05} = 1.96 \tag{1}$$

If the tests of kurtosis and skewness both prove $u < u_{0.05} = 1.96$, the data complies with normal distribution [6]. Table 1 showed that the kurtosis (g_1), skewness (g_2) and standard error σg_1 & σg_2 of the surface layer was 0.19, 0.046 and 0.097 & 0.049 respectively, and the u test values of kurtosis and skewness was 1.959 and 0.474 respectively, i.e., after eliminating the outliers, the data of the surface layer soil matches lognormal distribution; similarly, the kurtosis, skewness u test values of the sub-layer data after eliminating the outliers and logarithmic transformation was 0.231 and 0.826 respectively, i.e., the sub-layer data also matched lognormal distribution.

Theoretical semivariogram model fitting and analysis

Spatial analysis of surface layer As-concentrations. Descriptive statistical analysis can only roughly reflect the As-concentrations of the study area but not the structure and randomness, so geostatistical semivariograms were used to further explain the spatial structure and randomness of soil As-concentrations in the study area. Semivariogram maps can usually be fitted with some curve equations, which are called the theoretical semivariogram models. Theoretical semivariogram models may be spherical, Gaussian, Exponential, Linear and so on.

Determination of theoretical semivariogram models . On the geostatistical software GS +9.0 platform, different types of models were used to fit As semivariograms in isotropic soil and calculate the parameters, coefficient of determination (R^2) and residual sum of squares (RSS) of the models. The results are as shown in Table 2.

Table 2 Semivariogram model parameters for surface layer soil As-concentrations

Model	C_0 (nugget)	C_0+C (sill value)	A (range)	C_0/C_0+C	R^2 (coefficient of determination)	*RSS* (residual sum)
Exponential	0.598	1.298	280.16	0.458	0.913	0.0528
Spherical	0.820	1.044	302.25	0.785	0.880	0.0733
Gaussian	0.668	1.337	322.55	0.500	0.852	0.1100
Linear	0.837	1.475	344.93	0.568	0.866	0.0814

The semivariogram model of the maximum R^2 and minimum RSS is the optimal model [9-11]. In Table2, the coefficient of determination of the exponential model was 0.913, the largest in the four models, and its residual sum of squares was 0.0528, the smallest in the four models, so the optimal As-concentration semivariogram model of the soil in the study area was exponential.

Exponential model parameter analysis.(1) Semivariogram value at the origin is called nugget (C_0), usually indicating the variation caused by experimental sampling scale being larger than spatial scale, which is a kind of random variation. Random factors include, for example, fertilization, farming, planting, external sources of contamination and other human interference. Sill value usually indicates the total variation within the system, and it is the sum of structural variation and random variation. Structural factors include, for example, climate, parent material, topography, soil type, and so on. In theory, the heavy metal semivariogram nugget effect in soil is zero in natural conditions without human disturbances.

It can be learned from Table 2 that the exponential model nugget was 0.5957, showing a significant nugget effect, while the sill (C_0 + C) was 1.2983, indicating that the study area was evident subject to human disturbance apart from its own soil parent material. In addition, there may be sampling errors and random errors caused by determination and short distance variations. Combined with the actual situations of the study area, this human interference was originated from the As-containing high-hazard chemicals scattered into the soil when the discarded shells were destroyed and the local farming activities.

(2) The nugget to sill ratio ($C_0/(C_0 + C)$) is an important indicator reflecting the degree of spatial heterogeneity of regionalized variables, which indicates the proportion of the spatial heterogeneity caused by the random part accounting for in the total system variation and can be used as a measure of the spatial correlation degree of variables. In accordance with the grading standards of regionalized variable correlation degree, if the value is less than 25%, it indicates that the spatial variation is mainly structural and of high spatial correlation; if the value is between 25%

and 75%, it indicates moderate spatial correlation; if the value is greater than 75%, it indicates weak spatial correlation.

Table 2 showed that the nugget / sill ($C_0/(C_0 + C)$) value is 0.458, indicating a medium degree of spatial correlation. The spatial variance of soil arsenic in the study was due to internal factors (or structural factors, such as climate, topography, soil type, etc.) and also some external factors (or random factors, such as farming, fertilizing, planting, etc.) to some extent, and these external human activities weaken the spatial correlation.

(3) The range (A) is a spatial autocorrelation measure of random variables, and it is related to the interaction of processes in which the observations and sampling measures affect the soil properties. Within the range, the variables are spatial autocorrelated, or else not [12].

It can be seen from Table 2 that the effective range was 280.16 m, indicating that As-concentrations within 280.16 m from a sampling point were spatial correlated, or else not.

Analysis of soil As-concentration optimal spatial interpolation models. Selection and optimization of Kriging interpolation help further accurately and intuitively analyze the spatial distribution and variation characteristics of As-concentrations in the soil. After determining the optimal semivariogram, the geostatistical extension module of Arc Map was used for Ordinary Kriging, Simple Kriging and Universal Kriging spatial interpolations, and the cross-validation parameter values of interpolations in different Kriging models are as shown in Table 3.

Table 3 Cross-validation results of Kriging models for surface layer soil As-concentrations

Method （Method）	ME (Mean prediction error)	RMSE (Root mean square prediction error)	ASE (Mean standard error)	MSE (Standard mean error)	RMSSE (Standard root mean square prediction error)
Ordinary Kriging (OK)	2.864	54.06	94.94	0.0203	0.608
Simple Kriging (SK)	1.773	57.48	61.01	-0.0541	1.015
Universal Kriging (UK)	-2.172	54.95	69.89	-0.2187	1.439

According to the cross-validation parameter evaluation criteria generated by the Kriging output prediction surface: the interpolation method with ME and MSE close to 0, RMSE and ASE as close as possible and RMSSE close to 1 had relatively high accuracy and was thus the optimal interpolation model [13]. Comparing the interpolation results as shown in Table 3: ME and MSE of Simple Kriging interpolation method was 1.773 and -0.0541 respectively, closer to 0 than ordinary Kriging (2.864, 0.0203) and the Universal Kriging (-2.172, -0.2187); RMSE and MSE of Simple Kriging interpolation was 57.48 and 61.01 respectively, more closed to each other than Ordinary Kriging (54.06, 94.94) and Universal Kriging (54.95, 68.89); RMSSE of Simple Kriging interpolation was 1.015, more closer to 1 than 0.6081 and 1.439. In summary, the optimal spatial distribution interpolation method was Simple Kriging.

Spatial analysis of sub-layer As-concentrations. Correlation analysis of arsenic in the sub-layer soil was carried out by the same methods and procedures as for the surface layer and the results were compared.

Optimal models and parameters. The optimal semivariogram and Kriging interpolation for sub-layer As-concentrations were determined based on the descriptive statistic values.

Table 4 Semivariogram model parameters and Kriging models for the sub-layer As-concentrations

Analysis item	Optimal choice	Parameters					
Semivariogram model	Gaussian	C_0	C_0+C	A	RSS	R^2	C_0/C_0+C
		0.3570	0.7930	593.0	0.1430	0.6940	0.4500
Kriging interpolation	Simple Kriging	ME	RMSE	ASE	MSE	RMSSE	
		0.2019	37.35	36.95	-0.0442	1.106	

It can be seen from Table 4 that the optimal semivariogram for the spatial analysis of sub-layer As-concentrations in the study area was the Gaussian model, and the optimal Kriging interpolation was Simple Kriging.

Spatial interpolation of As-concentrations in the soil

Spatial interpolation of surface soil As-concentrations. The aforementioned optimal Kriging interpolation model was used to interpret and further analyse the surface As-concentrations in the study area. Simple Kriging interpolation was conducted in the geostatistical extension module of Arc Map to obtain the spatial distribution of arsenic in the soil in the study area (Figure 2).

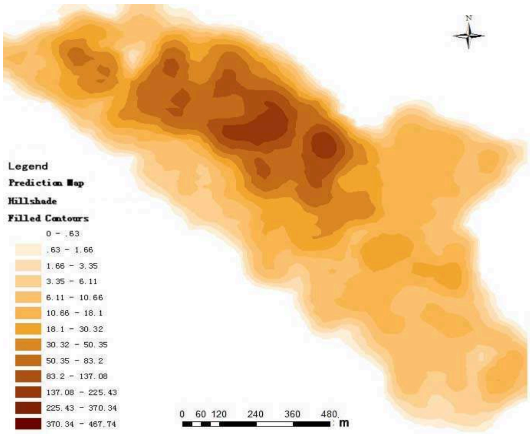

Fig.2 Map of Kriging estimates for surface layer soil As-concentrations

Figure 2 shows that the spatial distribution of arsenic has obvious regional characteristics.

(1) The high As-concentrations were distributed in the middle of the intersection of the farmland and the woodland in the study area and form a large area of high As-concentrations (30 ~ 300 $mg \cdot kg^{-1}$), accounting for 1/5 of the total study area. It can be speculated that the high As-concentration area was the central area where the discarded shells filled with high-hazard arsenic chemicals were destroyed or those not completely destroyed buried in centralization. The As-toxicants scattered due to the explosion or leaked during the burial time caused significantly higher As-concentrations in the area than surrounding areas.

(2) The As-concentrations were highest in the intersection of the farmland and the woodland and shown a decreasing trend in the directions of the northwest and southeast. This trend was, on the one hand, related to the layout of the shell destruction pits, and on the other hand, related to the geographical location of the study area - the study area was a SE-NW trending belt valley with hills on the south and north sides, and east and west wind might cause the As-containing high-hazard chemicals leaked due to corrosion and or released from destruction to contaminate the area and decline along the SE-NW direction. In addition, previously soybean was grown in the area, and the SE-NW plots and ridging mixed the arsenic contamination generated by the explosion and made the trend more apparent.

(3) There were small high As-concentration plaques in the northwest corner and the southeast corner. The reasons might be that: On the one hand, the northwest corner and the southeast corner

were near the road and easy to set up sporadic shell pits; on the other hand, the northwest corner was near the local resident blocks, and local residents might take the scattered shells filled with As-containing high-hazard chemicals and re-discarded or re-buried them.

Spatial interpolation of sub-layer As-concentrations. With the same coordinates for the surface, Simple Kriging interpolation was carried out for sub-layer As-concentrations, and the results are as shown in Figure 3.

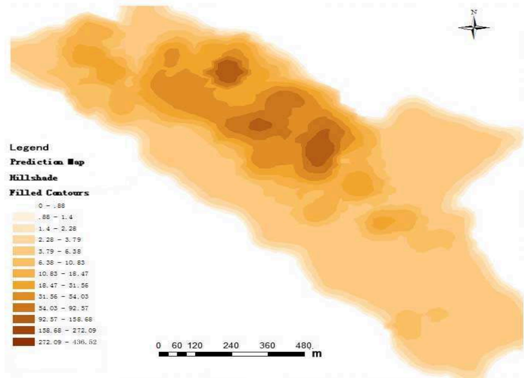

Fig.3 Map of Kriging estimates for the sub-layer soil As-concentrations

It can be seen from Figure 2 and 3 that the spatial distribution of the surface layer and that of the sub-layer were overall similar. The high As-concentrations were centralized in the intersection of the farmland and the woodland in the central part of the study area. The difference was that the As-concentrations of vast majority of the surface layer higher than the sub-layer except that at sampling points the As-concentrations were higher in the sub-layer than in the surface layer. The reasons for excessive arsenic in depth: (1)location of the shell destruction pits; (2) leakage of unexploded or incompletely exploded shells during burial; (3) disturbance of the contaminated soil during the digging process.

The surface As-concentrations (y) was weakly positively correlated to the sub-layer(x) ($y = 0.226x + 77.416$, $r = 0.231$), indicating that the surface layer was not obviously linear to the sub-layer.

Spatial analysis of soil profile As-concentrations. Soil profile samples were taken from a representative farmland area for analysis, of which the results are as shown in Figure 4.

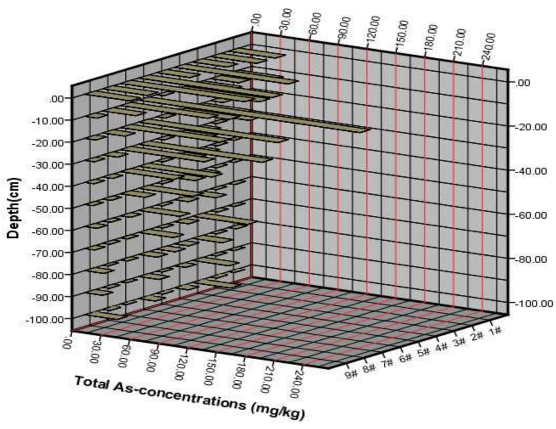

Fig.4 Map of soil profile As-concentrations in farmland

It can be seen from Figure 4 that the largest As-concentrations were basically on the surface (-20 cm ~ 0 cm); at most of the sampling points, and the As-concentrations decreased sharply as the soil depth increases and generally reached a very low level at 1.00 m from the surface, almost all below 5 $mg \cdot kg^{-1}$; almost all samples with excessive levels of As-concentrations were taken above -30 cm from the surface. There findings further suggested that, except the destruction pits, the soil contamination was caused by scattered As compounds from the explosion and that the depth of contamination was the result of combined effects of explosion scatters, human disturbance, natural migration of arsenic, and so on. The arsenic vertical distribution findings provided fundamental reference data for the restoration of the land farmland in the study area.

Conclusions

Based on the solid As-concentration findings of the survey, statistical methods were used to investigate the soil As-concentration semivariograms and the optimal fitting model, the GIS platform was used to compare and analyze multiple As-concentration spatial interpolation models, the optimal interpolation model was screened out and the soil As-concentration spatial distribution characteristics were obtained as follows:

(1) The As-concentration coefficient of variation in the surface layer (-0.50m ~ 0) in the study area reached 1.76, indicating obvious heterogeneity. The spatial variance was mainly subject to the effects of explosion scatters, leakage of buried unexploded shells filled with As-containing high-hazard chemicals, digging, crop planting and so on, showing a moderate degree of spatial correlation; the optimal semivariogram model was exponential, and the optimal Kriging interpolation model was Simple Kriging; the surface spatial interpolation results showed obvious regional characteristics.

(2) The As-concentrations in the soil sub-layer (-1.00 m ~ -0.50 m) were significantly reduced compared to that in the surface layer, and the sub-layer As-concentrations (y)was weakly positively correlated to that of the surface layer （x） ($y = 0.226x + 77.416$, $r = 0.231$). The optimal Semivariogram was the Gaussian model, the optimal interpolation model was Simple Kriging, and the two were of similar spatial structures.

(3) The soil profile analysis showed that the arsenic contamination was mainly concentrated in the surface soil, which decreased as the soil depth increases in most of the sampling points.

This paper studied the arsenic spatial structure and spatial distribution characteristics in the study area; Simple Kriging interpolation was used to obtain the arsenic spatial distribution map, which helps simply, intuitively and comprehensively understand the spatial variance and distribution of arsenic in the study area and has certain significance to the treatment and restoration of the study area.

References

[1] C.B. Zhang, Z.B. Li, C.X. Yao, et al. Characteristics of spatial variability of soil heavy contentsin contaminated sites and their implications for source identification. Soils. 8(2006) 525–533.

[2] N.Z. Lei , X.Y. Wang, J.G. Jiang Comparison of Kriging interpolation precisions in different topographic units and number of samples: a case study of spatial distribution of soil total nitrogen in the Shucheng County. Hydrogeology & Engineering Geology. (2008) 86–91.

[3] C.Z. Wang, Y.K. Peng, S.Q. Liu. Survey on forest soil in Dunhua forest-region, Jilin, China. Jilin Forestry Scicnce and Technology.32(2003) 49–50.

[4] Z.Q. Wang. Geostatistics and its application in the studies of soil ecology. Science Press Ltd, Beijing.1999.

[5] L. Chen, J.F. Zhao. Enviromental monitoring. Chemical Industry Press, Beijing.2004.

[6] J.H. Xu. Quantity geography. Higher Education Press, Beijing. 2006.

[7] Q.W. Liu, Z.Y. Hu. How to do normality test with SPSS and SAS statistical software. Journal of Xiangnan University (Natural Sciences). 7(2005) 56–58.

[8] J.Q. Xia, Y.M. Luo. Several key issues in research of soil environmental quality in china. Journal of Ecology and Rural Environment.23 (2007) 1–6.

[9] H.L. Zheng, J. Chen, W.J. Deng, et al. Assessment of soil heavy metals pollution in the chemical industrial areas of Nanjing peri-urban zone. Acta Scientiae Circumstantiae. 25(2005)1182-1188.

[10] N.M. Zhang, B.G. Li, K.L. Hu. Spatial variabilities of heavy metals and salt contents in soils of Taiyuan sewage irrigation region. Acta Scientiae Circumstantiae. 21(2001) 349 – 353.

[11] Y.C. Wei, S.Z. Chen. Geogrphical modeling principle and method. Science Press, Beijing. 2004.

[12] X.Y. Li, S.W. Zhang, Z.M. Wang, et.al. Spatial variability and pattern analysis of soil properties in Dehui city, Jilin province. Journal Of Geographical Sciences, 59.(2004).

[13] K.L. Hu, B.G. Li, Y.Z. Lv. Comparison of various spatial interpolation methods for non-stationary regional soil mercury content. Chinese Journal Of Environmental Science. 25(2004) 132 – 137.